

## 2.2 Project Planning and Sourcing Data with an API

The strategic dashboard for Citi Bike should begin with an executive summary highlighting the analysis goal of diagnosing bike and dock shortages and recommending rebalancing actions, emphasizing that the top 20-30 stations handle most of the demand. It should also note the exclusion of bike type and user-type metrics for focusing on operational efficiency. Key Performance Indicators (KPIs) such as Bike Shortage Frequency, Dock Shortage Frequency, and a Rebalancing Index should be prominently displayed. The temporal analysis will focus on identifying when shortages occur, visualizing hourly and daily trends with interactive filters, and highlighting seasonal patterns with heatmaps. Geospatial analysis should pinpoint where these shortages are concentrated, focusing on the top stations with a shortage heatmap overlaid on an NYC map, allowing drill-down to individual station details. Actionable recommendations should provide specific rebalancing strategies, suggesting routes and times for bike redistribution based on the Rebalancing Index, and propose capacity changes like increasing docks at stations with frequent shortages. Interactive features should include station and time range selection, along with a shortage threshold setting. This approach ensures a highly focused and actionable dashboard for diagnosing and addressing bike and dock shortages within the Citi Bike system.

### **Where are bike and dock shortages most frequent?**

*Visualisation:* Heatmap overlaid on a map of NYC, showing the frequency of both bike shortages (stations running out of bikes) and dock shortages (stations being full) at each station. Different color scales or shading could be used to distinguish between bike and dock shortages, or a combined "shortage index" could be represented by a single color scale.

*Rationale:* Provides a clear visual representation of geographic hotspots for both bike and dock shortages in a single view, allowing for quick identification of problem areas.

### **When (what time of day/day of week) are bike shortages most common?**

*Visualisation:* Heatmap with hours of the day on the vertical axis and days of the week on the horizontal axis. Each cell will be coloured based on the frequency of bike shortages during that time period, with darker colors indicating more frequent shortages.

*Rationale:* Identifies the most critical times of day and days of the week to focus rebalancing efforts.

### **When (what time of day/day of week) are dock shortages most common?**

*Visualisation:* Heatmap similar to question 2, but showing the frequency of dock shortages.

*Rationale:* Identifies the most critical times of day and days of the week when stations are frequently full and unable to accept returned bikes.

**How do the frequency of bike/dock shortages change throughout the year?**

*Visualisation:* Line chart with months on the horizontal axis and shortage frequency (separate lines for bike and dock shortages) on the vertical axis.

*Rationale:* Reveals seasonal trends in bike and dock availability, helping to anticipate and prepare for peak and off-peak periods.

**How can weather conditions influence shortages?**

*Visualisation:* Scatter plots showing relationships between bike/dock shortages and temperature, precipitation, etc.

*Rationale:* Reveals if weather conditions drive patterns of high/low demand.

**Can we find any co-relation between bike shortages and dock shortages. e.g if bikes are used more at some place are they being returned or not?**

*Visualisation:* scatter plot will plot shortage of bikes vs shortage of docks. We'd plot each station as a distinct datapoint.

*Rationale:* Correlation between the metrics for each data point will determine if the shortages are linked to each other.